This listing of claims will replace all prior versions, and listings, of claims in the application:

Claims 1-45 (deleted)

Claim 46 (currently amended): A method for filtering a plurality of candidate search results to remove near-duplicates, the method comprising:

a) for ~~each of a plurality of~~ one of the plurality of candidate search results, determining whether the one candidate search result is a near-duplicate of another of the plurality of candidate search results ~~result~~ by

1) comparing a cluster identifier of the one candidate search result with ~~that~~ a cluster identifier of the other candidate search result, and

2) if the cluster identifiers of the one and the other ~~two~~ candidate search results match, then concluding that the one ~~two~~ candidate search is a near-duplicate of the other candidate search result ~~results are near-duplicates~~; and

b) ~~if it is determined~~ in response to a determination that the one candidate search result is a near-duplicate of the other candidate search result, ~~then~~ rejecting the one candidate search result thereby defining a filtered set of search results including only those of the plurality of candidate search results that have not been rejected.

Claim 47 (currently amended): A search filter for processing a plurality of search results to remove near-duplicates, the search filter comprising:

a) a near-duplicate determination facility for determining, for ~~each of a plurality of~~ one of the plurality of candidate search results, whether the one candidate search result is a near-duplicate of another of the plurality of candidate search results ~~result~~, and wherein the near-duplicate determination facility includes a comparison facility for comparing a cluster identifier of the one candidate search result with a cluster identifier ~~that~~ of ~~another~~ the other candidate search result, and wherein if the cluster identifiers of the one ~~two~~ candidate search ~~results~~ result and the other candidate search result match, then it is concluded that the one ~~two~~ candidate search ~~results~~ result and the other candidate search result are near-duplicates; and

b) a filter for rejecting the one candidate search result if it is determined that the one candidate search result is a near-duplicate of the other candidate search result and passing the one candidate search result if it is not determined that the one candidate search result is a near-duplicate of the other candidate search result, thereby defining a filtered set of search results including only those of the plurality of candidate results that have not been rejected by the filter.

Claim 48 (previously presented): A machine-readable medium having stored thereon a plurality of records, each of the records comprising:

a) a first field for storing a document identifier; and

b) a plurality of lists, each of the plurality of lists containing elements of a document identified by the document identifier stored in the first field,

wherein a hash function is used to hash each of the elements in order to determine which one of the plurality of lists that each of the elements will be contained in.

Claim 49 (currently amended): A method for determining whether two documents are near-duplicates, the method comprising:

a) for each of the two documents, generating at least two <u>different</u> fingerprints; and

b) determining whether or not the two documents are near-duplicate documents by

1) determining whether or not any one of the at least two fingerprints of a first of the two documents matches any one of the at least two fingerprints of a second of the two documents, and

2) if it is determined that any one fingerprint of the at least two fingerprints of the first of the two documents does match any one fingerprint of the at least two fingerprints of the second of the two documents, then concluding that the two documents are near-duplicates<u>; and</u>

<u>c) using the determination of whether or not the two documents are near-duplicates in at least one of (A) an act of serving search results corresponding to documents, (B) an act of</u>

crawling documents, (C) an act of indexing documents, and (D) an act of fixing a broken link to at least one of the two documents.

Claim 50 (previously presented): A machine-readable medium having stored thereon a plurality of records, each of the records comprising:

a) a first field for storing a document identifier; and

b) a plurality of lists, each of the plurality of lists containing elements of a document identified by the document identifier stored in the first field,

wherein at least some of the plurality of lists include different numbers of elements.

Claim 51 (currently amended) The machine-readable medium of claim ~~51~~ 50 wherein at least one of the plurality of lists include no elements.

Claim 52 (currently amended): A machine-readable medium having stored thereon a plurality of records, each of the records comprising:

a) a first field for storing a document identifier; and

b) a plurality of lists, each of the plurality of lists containing elements of a document identified by the document identifier stored in the first field,

wherein at least some contiguous elements in a document are not ~~necessarily~~ contiguous elements of a list.

Claim 53 (previously presented): A machine-readable medium having stored thereon a plurality of records, each of the records comprising:

a) a first field for storing a document identifier; and

b) a plurality of lists, each of the plurality of lists containing elements of a document identified by the document identifier stored in the first field,

wherein for each of the records, the number of lists is the same.

Claim 54 (previously presented): The machine-readable medium of claim 53 wherein a number of the plurality of lists is independent of document size.

Claim 55 (previously presented): The machine-readable medium of claim 48 wherein each of the elements of a document is an element that has been extracted from the document.

Claim 56 (previously presented): The machine-readable medium of claim 48 wherein each of the elements of a document is a predetermined one of (A) a predetermined number of words, (B) a predetermined number of sentences, (C) a predetermined number of characters, (D) a predetermined number of paragraphs, and (E) a predetermined number of sections.

Claim 57 (previously presented): The machine-readable medium of claim 48 wherein each of the elements of a document partially overlaps another of the elements of the document.

Claim 58 (previously presented): The machine-readable medium of claim 50 wherein each of the elements of a document is an element that has been extracted from the document.

Claim 59 (previously presented): The machine-readable medium of claim 50 wherein each of the elements of a document is a predetermined one of (A) a predetermined number of words, (B) a predetermined number of sentences, (C) a predetermined number of characters, (D) a predetermined number of paragraphs, and (E) a predetermined number of sections.

Claim 60 (previously presented): The machine-readable medium of claim 50 wherein each of the elements of a document partially overlaps another of the elements of the document.

Claim 61 (previously presented): The machine-readable medium of claim 52 wherein each of the elements of a document is an element that has been extracted from the document.

Claim 62 (previously presented): The machine-readable medium of claim 52 wherein each of the elements of a document is a predetermined one of (A) a predetermined number of words, (B) a predetermined number of sentences, (C) a predetermined number of characters, (D) a predetermined number of paragraphs, and (E) a predetermined number of sections.

Claim 63 (previously presented): The machine-readable medium of claim 52 wherein each of the elements of a document partially overlaps another of the elements of the document.

Claim 64 (previously presented): The machine-readable medium of claim 53 wherein each of the elements of a document is an element that has been extracted from the document.

Claim 65 (previously presented): The machine-readable medium of claim 53 wherein each of the elements of a document is a predetermined one of (A) a predetermined number of words, (B) a predetermined number of sentences, (C) a predetermined number of characters, (D) a predetermined number of paragraphs, and (E) a predetermined number of sections.

Claim 66 (previously presented): The machine-readable medium of claim 53 wherein each of the elements of a document partially overlaps another of the elements of the document.

Claim 67 (previously presented): A machine-readable medium having stored thereon a plurality of records, each of the records comprising:

a) a first field for storing a document identifier; and

b) a plurality of lists, each of the plurality of lists containing elements of a document identified by the document identifier stored in the first field, wherein each of the elements are contained in one of

the plurality of lists in accordance with a result of hashing the element using a hash function.